Computer Science Department

TECHNICAL REPORT

A Meaning Representation
for Generic Sentences

T. Strzalkowski

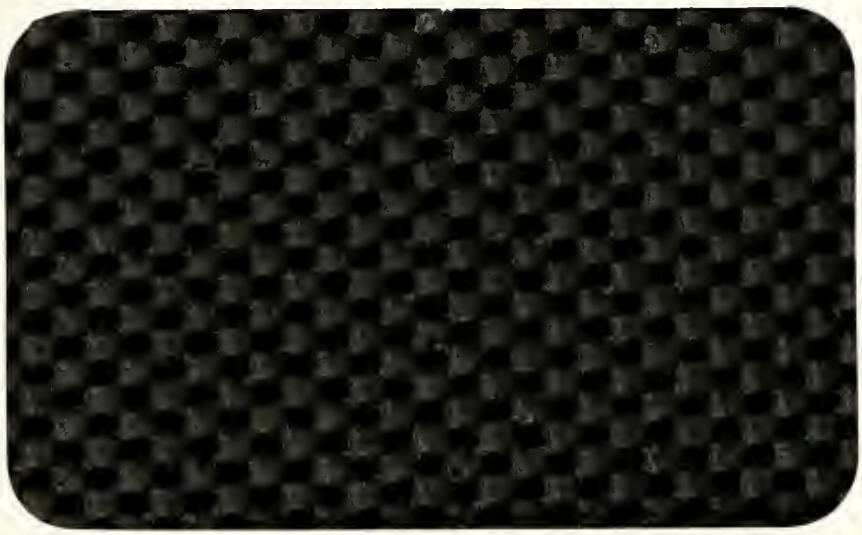
Technical Report 423

December 1988

NEW YORK UNIVERSITY

Department of Computer Science
Courant Institute of Mathematical Sciences
251 MERCER STREET, NEW YORK, N.Y. 10012

NYU COMPSCI TR-423
Strzalkowski, Tomek
A meaning representation
for generic sentences.
c.1



A Meaning Representation for Generic Sentences

T. Strzalkowski

Technical Report 423

December 1988

A Meaning Representation for Generic Sentences

Tomek Strzalkowski
Courant Institute of Mathematical Sciences
New York University
715 Broadway
New York, NY 10003
U.S.A.

Abstract

The subject of this paper is an adequate meaning representation for a certain type of generic sentences in natural language. Specifically, the following problems are addressed. (1) What does it mean for a sentence to be generic and what sets it apart from other non-singular sentences, (2) what kind of representation do we need to capture the meaning of a generic sentence, together with its truth conditions and, possibly, some implications.

1. Introduction

The subject of this paper is an adequate meaning representation for generic sentences in natural language, such as these listed below.

(1) Birds fly.
(2) Kenyans are good Olympic runners.
(3) Americans drive big cars.
(4) Physicists win Nobel Prizes.
(5) Germans are taller than Italians.

Generic and other non-singular sentences have been the subject of much debate in philosophy, linguistics and computer science, see (Quine 1960), (Kripke 1972), (Montague 1974), (Carlson 1982), (Sidner 1979). The major problem, it seems, was in capturing the exact truth conditions of such sentences. Let us take, for example, (1). Does it mean that *all* birds fly, in other words, that

(6) $\forall x[\text{bird}(x) \rightarrow \text{flies}(x)]$

is true? Certainly not, because a bird with a broken wing does not fly, nor does a penguin. A natural reaction to this challenge is to relax the strict universal quantification to something a little weaker, like *most*. Thus, (1) is taken to mean that *most* birds fly. Unfortunately, *most* is itself ambiguous; does it mean more than a half, at least two thirds, 99 percent? The following pairs of sentences indicate that although *most* may sometimes accurately reflect the meaning of a generic statement, it is not always so.

(7) Most NYU Computer Science graduate students are foreigners. (*true*)
NYU Computer Science graduate students are foreigners. (*false*)
(8) Most lotery ticket buyers loose. (*true*)
Lotery ticket buyers loose. (*false* ?)
(9) Most Americans voted republican. (*true*)
Americans voted republican. (*true* ?)
(10) Most presidents live in the White House. (*false*, too weak)
Presidents live in the White House. (*true*)

It appears that we need something stronger and less ambiguous, like *nearly all*, or *all except for a negligible amount of exceptions*. [1] As soon as we strengthen *most* to a nearly universal quantification, however, another problem pops up. In this new interpretation (1) comes forth as a false statement, contrary to our intuition. What seems to be missing is a context: surely we do not mean here that nearly all birds fly, but perhaps, that nearly all birds that are so and

[1] One may be tempted to use *normally*, or perhaps *usually*. However, this in turn requires capturing the semantics of these words, which appear to be quantitatively based as well.

so fly (for example, wild birds in moderate climate zone, but arguably, not in arctic climate zone). The same kind of analysis can be applied to (2). Thus, *Kenyans are good Olympic runners* is interpreted as *nearly all Kenyans who are Olympic runners are good Olympic runners*. Of course, we may obtain different interpretations as we change the context in which the *nearly all* quantifier is defined. Let us look at (3) and (4) next. Can we find similar restrictive contexts in which nearly all Americans that are so and so drive big cars, and nearly all physicists that are this and that win Nobel Prizes without reducing these statements to banalities? In the case of (3) the context might be taken to include the Americans who drive cars, but in (4) there seem to be no such alternative, unless the preceding discourse furnishes one. If we are not willing to accept the context provided by the discourse at hand, which in the absence of other information contains all Americans (or all physicists), then the problem turns into a nightmare. Based on examples like these, and others, Carlson (1982) rejected the quantificational approach to representing generics as being entirely misdirected, see also (Schubert & Pelletier 1987) for an overview. We shall argue here that the quantificational approach can work quite well for a substantial subset of generic sentences. We point out that much of the confusion surrounding the semantics of generic sentences can be cleared away once we realize that they can assume one of the two possible interpretations: transparent or opaque. We suggest that the semantic distinction between these two interpretations can be systematically captured by introducing two new quantifiers into predicate calculus. To avoid confusion with other types of generic sentences we shall refer to them with an all-inclusive term of non-singular sentences. The first observation is that sentences (3) and (4) are ambiguous between their nearly universal reading (possibly in some restrictive context) and what appears to be their intended reading. In the nearly universal interpretation (3) means that nearly all Americans (perhaps those who drive cars) drive big cars, while (4) means that nearly all physicists (again, possibly in some context) win Nobel Prizes. These are the interpretations that are most frequently assumed, especially in the absence of any external context, and we propose to call these transparently generic. Nonetheless, we feel the primary readings of both (3) and (4) are not transparently generic. What (3) appears to mean is that, judging by some international standards (possibly in some context), Americans drive a *significant* or *non-negligible* portion of all big cars. Similarly, in (4) we may say that among all Nobel Prize winners (past, present and future) physicists create a significant or non-negligible group.[2] These are the interpretations that we call opaquely generic. In this paper we show that these two readings can and should be systematically separated, and that the notion such as *significant* or *non-negligible* amount can be given a precise interpretation.

The major objectives of this research can be stated as follows:

- What does it mean for a sentence to be generic? What sets a generic sentence apart from other non-singular sentences?
- How can one determine whether a generic sentence is true or false, and what are immediate consequences of such a sentence being true or false?
- What kind of representation do we need to capture the meaning of a generic sentence, together with its truth conditions and implications?

Before we address these issues directly, we define a few underlying concepts that would help to set this discussion in a proper context.

2. Non-singular terms in natural language

The concept of a non-singular term is best illustrated by the class of inter-sentential dependencies whereby a reference is made across boundaries of different detail-levels (or reference levels) in discourse (Strzalkowski 1986a-c), (Strzalkowski & Cercone 1986), (Strzalkowski & Cercone 1988). For example, in

(11) *John's new pet is an alligator. But the alligator cannot live in our climate.*

the alligator in the second sentence most likely refers to a generic concept of which the alligator in the first sentence is an instance or extension. Thus we can say that the second alligator is a non-singular *superobject* in which the first alligator somehow participates. The extent of such participation is not clear, but in general it can be observed that certain predications true of complexes of different kind are not preserved for their parts or elements, and vice-versa. To represent this kind of inter-sentential dependency we introduced[3] a multi-level model for interpreting the natural language expressions, such that the levels in the model would correspond (roughly) to the levels of reference

[2] We do not ignore the habitual reading of (4) *every now and then some physicists win some Nobel Prizes*. We shall argue, however, that the habitual interpretation alone does not accurately capture the meaning of this sentence.

[3] For more detailed presentation of this approach the reader is referred to (Strzalkowski 1986bc) and (Strzalkowski&Cercone 1988)

in discourse. For instance, in (11) the resulting representation would have both alligators placed at different, though related, "object levels". The levels are partially ordered with the *lower than* (i.e., *more detailed than*) relation with respect to some current level (corresponding to the reference level at a present point in discourse). The basic concepts within this approach are the notions of a singular object and a coordinate. A singular object is any entity that can be taken as a coherent whole, in other words, it can be referred to directly using a referring expression of language: a name, a definite description, a pronoun. A coordinate is usually an ordered set specifying a type of "dimension" that a singular object spans. The most common of these coordinates are time and space but other more abstract ones are also possible. These two basic concepts are then used to define the notion of the object's *instance* with respect to some coordinate. Thus, the pet alligator in the example above is related to the generic concept of alligator by some "species" coordinate that somehow enumerates all alligators around the world: the pet alligator becomes a lower-level instance of the higher-level "generic" alligator. If we reverse this process we can combine lower-level objects into higher-level complexes (or superobjects) to which we can subsequently refer using *collective* terms, singular or plural, such as, for example, *people* or *the man* (generic). The *lower than* relation between levels in the universe model derives from expanding the notion of instance over collections of objects. An important point here is that a coordinate defines a decomposition of a compound object into instances, and that an object may be decomposed with respect to different coordinates, yielding different sets of instances. For example, the abstract object corresponding to the generic term *birds*, as in *birds fly*, can be decomposed with the *genus* coordinate yielding such instances as *eagle*, *canary*, or *penguin*. It can also be decomposed with the *specimen* coordinate that would allow us to pick up single birds, such as Tweety the canary. Generic sentences of the type that we discuss in this paper are naturally understood in terms of decompositions being applied to higher-level objects.

3. Generic and other non-singular sentences

Although the statement *birds fly* is normally understood at some higher reference level, it does appear to convey some information about the concept's instances. In a default theory the statement would be taken to mean that given any instance of the concept B, referred to by *birds*, we shall assume that it flies, unless the contrary is known. This interpretation does not appear quite accurate, however. One persistent problem with nonmonotonic representation for generic sentences has to do with the ability to maintain a consistent set of beliefs. Although one can admit a number of exceptions while still believing that birds fly, this belief may eventually change if enough evidence is persistently provided. This is not so with default and nonmonotonic logics based on the notions of domain closure or predicate circumscription (Reiter 1980, McCarthy 1980). In particular, we do not have a way of determining when a generic statement can be taken as a truth and when it cannot, as is in the case with *penguins fly*. A more careful representation is required. Roughly, we would like to build an interpretation along the following lines. *Birds fly* is true relative to a coordinate *T* iff *T* is decomposing the generic object referred to by *birds* into some lower-level instances in such a way that except for a "negligible" number of exceptions, every instance in this decomposition has the property *flies*. In one possible interpretation, a number of exceptional instances of a concept is called "negligible" if the ratio of this number to the number of all instances of the concept approaches 0 when the number of all instances of the concept approaches infinity. Thus, if we agree that there is no reason to believe that the kind of penguin has at most negligibly many instances, then *birds fly* is false if the decomposition of *birds* is made with respect to the specimen coordinate (i.e., the one which picks up single birds). The statement is true, however, with respect to a different decomposition of *birds*, namely with the coordinate genus, under the assumption that only finitely many elements of this coordinate pick up classes of birds of which we can say that they do not fly. Another possibility emerges when we consider the statement as addressing only a restricted subclass of all birds, and therefore being interpreted in some context: *birds that are so and so fly*. This restrictive context may be provided by a current discourse situation, or it may come from the general knowledge about birds. The coordinate used to define lower-level instances of *birds* remains the same, only now the quantification is narrowed to a subset of it. According to the well-known conversational postulates, the context is normally understood as not being defined directly in terms of the sentence's main predicate, in this case *fly*.[4] From here we may, perhaps, go a little farther saying that a higher-level generic statement, like *birds fly*, is true (verifiable at some lower-level) in context C iff there exists a coordinate T such that the statement is true relative to $T \cap C$.

As indicated in the introduction, the above analysis constitutes just one possible interpretation of a non-singular sentence. Let us call this interpretation *transparently generic*, and let us call a sentence (or utterance)

[4] Another way of looking at things is to have the context-bound generic term refer to an object at some intermediate level between that of unrestricted generic term and that of its terminal instances.

transparently generic if its intended interpretation is transparently generic. Let S be a sentence containing a predicate P with an argument N which refers to a higher-level (non-singular) object ξ, and let P(N) be the translation of S at the corresponding reference level (for example, *fly(birds)*). Then S has a transparently generic interpretation only if there exists a decomposition of the object ξ with respect to some coordinate T, and such that for every instance ξ_t of ξ, for any $t \in T$, if N_t is a description that refers to the object ξ_t then $P(N_t)$[5] is well-defined. Some examples of such decompositions are shown in (12)-(15) below. Sentences in (a) are interpreted at higher levels, while those in (b) are interpreted at instance levels. In (12), (13) and (14) the decomposition is well-defined; in (15) it is not (coordinates included in parentheses).

(12) (a) Birds fly.
(b) A bird flies. (*specimen*)
(13) (a) Physicists win Nobel Prizes.
(b) Some physicists win some Nobel Prizes. (*specimen/groups*)
(14) (a) Presidents are elected every four years.
(b) The U.S. President is elected every four years. (*country*)
(15) (a) U.S. Presidents are elected every four years.
(b) The U.S. President from 1976 to 1980 is elected every four years. (*4-year intervals*)

The decomposition serves as the basis for determining the sentence's truth value. In case no well-defined decomposition into a lower level can be found, the truth conditions must be determined within the statement's own level, and no generic interpretation, as we understand it here, will be possible. Examples of such non-generic non-singular sentences involve predications referring to properties of higher-level objects which are not derived from the properties of their instances, such as *being numerous, widespread, frequent, scarce,* or even *extinct*. Thus, *birds are numerous, good Olympic runners are scarce*, or *presidents are elected every four years* (as in (16) but not as in (15)) are not considered generic. These sentences cannot be interpreted quantitatively with respect to the properties their instances may have.

Specifying a decomposition into a lower level is not a sufficient condition for a sentence to have a generic interpretation, however. Consider, for example, the sentence *Germans are taller than Italians*. A generic interpretation of this statement would require that when comparing a German and an Italian, in nearly all cases the German is taller than the Italian.[6] To put it into more formal set-theoretical terms, we take a Cartesian product of the sets of all Germans and all Italians, i.e., $GI=\{(g,i) \mid g \in \text{Germans} \ \& \ i \in \text{Italians}\}$. Except for a negligible number of exceptions, elements of GI are such that g is taller than i. [7] There are, however, other possible interpretations of this sentence which are not generic. One commonly used interpretation involves the statistical notion of "average" element of a set; an average German is taller than an average Italian. Statistical interpretations may approximate the generic interpretations to some degree, but in the worst case they have little to do with single instances and should be interpreted at the higher level only. Consider, for example, the interpretations where two sets A and B are being compared by matching each member from one set with one member from the other set. The result of such a match could be either win or lose. The set whose elements collect more wins is considered better, and we are justified to say, for instance, that A's are better than B's. The statement is not generic, though; we do not know enough about members of both sets to make a truthful generic claim. The comparison is made between non-singular objects A and B, and the interpretation can only make sense at the level these objects belong to.

---

[5] P may need to be adjusted for tense and/or plurality of object NP's (if any).

[6] We assume here, for simplicity, that both Germans and Italians constitute homogeneous adult populations. In practice, these sets are subdivided into age groups, etc., and comparison is made between corresponding groups only. This subdivision may be the source of the restrictive context in which the sentence is understood (for example, *adult Germans in the age group 28-35 are taller than Italians in the same age group*). It is also important that the compared sets be of comparable cardinality, i.e., one set cannot be negligibly small as compared to the other. A global statement about the entire populations may be the result of an arbitrary combination of the age-group comparisons, and thus may not be generic with respect to individuals.

[7] In a recent paper, Scha and Stallard (1988) suggest an interpretation that would go along the following lines: for (nearly) every German there is an Italian that is shorter than him, and for (nearly) every Italian there is a German that is taller than him. Although the examples they consider involve only universal quantification over finite sets, they claim that such interpretation would work even in the presence of exceptions, what can be achieved by replacing *every* by *nearly every* in the above. The resulting interpretation is unacceptable, however, because one can easily produce situations where both *Germans are taller than Italians* and *Italians are taller than Germans* is true at the same time, despite the fact that only one of them, and sometimes none, can be so regarded.

Let us consider now a different example, repeated here from section 1.

(4) Physicists win Nobel Prizes.

Is this sentence generic or not? It establishes a meaningful decomposition of the non-singular object denoted by *physicists* into individuals, each of whom can be said, truthfully or not, to win some Nobel Prizes. [8] And yet the transparently generic interpretation goes counter to our intuition: it is not true that *nearly all physicists win Nobel Prizes*. It is pointless to look for some restricting context that would allow us to understand this statements in some narrower sense. There is one other possibility left, though. Instead of looking at the set of physicists, we turn our attention to the set of individuals who win Nobel Prizes. How many physicists are among them, do they create a significant, non-negligible group? If so, then our claim that physicists win Nobel Prizes is justified, provided that they will remain a non-negligible group as the set of Nobel Prize winners is expanded. When presented with this sentence, native speakers tend to select the habitual reading that goes *every now and then some physicists win some Nobel Prizes*, and reject the reading argued for above. However, when challenged as to the exact character of the repetition implied by the habitual interpretation they agree that it must be at a pace with the number of prizes being awarded, and that physicists must remain a visible group among other Nobel Prize winners. The point is strengthened if we consider the passive version of (4).

(4') Nobel Prizes are won by physicists.

Now, (4') is normally taken to mean that a substantial, or non-negligible, portion of Nobel Prizes are being won be physicists. [9] Thus we arrive at another decompositional interpretation of non-singular sentences, which we would like to call the *opaquely generic* interpretation, because of a certain similarity between this interpretation and opaque readings of some class of universal statements (see, for example, Partee 1978).[10] Therefore, a non-singular sentence N(P) is true in its opaquely generic interpretation just in the case when, among all lower level objects that have the property N (i.e., Nobel Prize winners) and are of the same singularity as the instances of the object denoted by P, the instances of the object denoted by P (i.e., single physicists or groups of physicists) create a non-negligible subclass. In other words, as we grow the set N, its intersection with P will grow at a comparable pace.

An interesting thing to note is that generic sentences usually assume either transparent or opaque reading as the preferred one. There seem to be situations, however, where both interpretations are equally possible, for example *birds fly*: either nearly all birds fly (in some context), or birds are non-negligible among those vertebrates (all animals, things, etc.) that fly. This may be the reason why the transparently generic interpretation of this sentence does not always appear satisfactory.

The next thing to look at are habitual statements, which may be considered as a special kind of generic statements where the time coordinate (in various guises) plays a leading role. We will not have much to say about these sentences, except that they display some similarities to other generic sentences. Thus, statements like *John smokes*, *John walks to work*, *This department handles mail arriving from Antarctica*, or *Mary writes to her mother*, describe activities which involve an element of repetition over time. Of course, this "repetition" has a different character for each of these statements. In the case of *John smokes*, for example, we would normally require that the activity occurs at least once a day, and therefore we may introduce a *days* coordinate and verify the statement relative to it. Note that a verification with respect to "hours" coordinate may render this statement false, which would be appropriate if, according to some arbitrary standards, smoking once a day is not considered as smoking. As we may expect, the coordinate we select for verification of a habitual statement, although context-dependent, cannot be an arbitrary one. The coordinate must support the repetitive character of the habitual statement over some temporal context Θ. If we assume an interval-based time coordinate T, such that $\cup_{t \in T} t = \Theta$ and $\cap_{t \in T} t = \emptyset$, then the largest element of T (that is, the longest interval t of T) is no larger than a certain preset, contextually bound value. In the absence of such context we may assume that the largest interval in T is finite. For example, for *Mary writes to her mother* the habitual interpretation requires that *Mary is writing to her mother* is repeated at least once a month (2 months, a year, etc.).

---

[8] The predicate *win Nobel Prizes* takes one of the two possible decompositions, as we decompose physicists into lower-level instances (individuals or groups). It may be either *win(s) Nobel Prizes*, a habitual, or *win(s) some Nobel Prize(s)*. We assume the second of these possibilities here. Habituals are discussed later in this section.

[9] A number of informants have agreed that this is the preferred interpretation.

[10] These similarities have chiefly to do with the quantifier scoping.

For *John walks to work* the situation is somewhat more complicated. In one reading, when walking to work is the activity described, the lower-level interpretation is similar to that of *John smokes*. In another reading, the sentence is understood more like *John walks when he goes to work*, in which case we have two different activities described: going to work, and walking while going to work, where the second activity is relative to the first. Interpretation of the first of these activities will determine an appropriate time coordinate which can be used in interpretation of the second activity. It follows that the first activity (going to work) must be habitual in order for the second activity (walking to work) to be habitual. The interpretation we look for is this: the statement *John walks to work* is true (and we do not have to add "relative" this time because this relativity is made explicit) iff except for a "negligible" number of exceptions, whenever John goes to work, he walks.[11]

This is only half of the story, however. Let us look at *John smokes* again, but this time replacing John by some unknown description X, i.e., *X smokes*. Except for the unrestricted habitual reading as discussed above, there is another, restricted reading, as in *This machine smokes [whenever used]*. Here, again, we quantify over the instances of the restricting event *X is being used*, whose actual occurrence is irrelevant to the truth of the original statement. Note that in this interpretation, the statement is not a habitual but an ordinary singular statement, much like *X is red* or *X can fly*, in other words, merely expressing X's membership is a class of things of which we know that they smoke, are red, or can fly. Habitual statements, on the other hand, are so understood only by virtue of their possible interpretation at some lower level and such that the decomposition involves a time-based coordinate.

4. Meaning representation for generic sentences

We propose that two new quantifiers be introduced into the predicate calculus. These are $\overset{\circ}{\forall}$ and $\overset{\circ}{\exists}$ which should be read, respectively, as *all except for a negligible number of exceptions*, and *a non-negligible amount*. These quantifiers are well-defined only if the immediate context on which they operate is known. Thus, it is more appropriate to consider these symbols together with a context-setting set, that is $\overset{\circ}{\forall}\{x\in P\}$. For example, the transparently generic representation for *birds fly* would be

$$\overset{\circ}{\forall}\{x\in \text{Birds}\}[\text{fly}(x)]$$

or, more appropriately,

$$\exists T_B \overset{\circ}{\forall}\{t\in T_B\}[\text{fly}(B_t)]$$

where B refers to the higher-level object denoted by *birds*, and T_B is a coordinate decomposing B. Similarly, *physicists win Nobel Prizes* receives the opaquely generic interpretation as

$$\overset{\circ}{\exists}\{x\in \text{Win-NP}\}[\text{physicist}(x)]$$

or, perhaps,

$$\overset{\circ}{\exists}\{y\in \text{NP}\}\exists x[\text{physicist}(x)\ \&\ \text{wins}(x,y)]$$

where NP denotes the higher-level concept of *Nobel Prizes*. This formula can be rewritten in a more appropriate form, with the use of coordinates, as

$$\exists T_{NP}\exists T_{Ph}\overset{\circ}{\exists}\{t\in T_{NP}\}\exists\{s\in T_{Ph}\}[\text{wins}(Ph_s,NP_t)]$$

Although the multi-level representation is more appropriate,[12] we shall use the simplified single-level version throughout this section to make some important properties of the new quantifiers transparent.

Of course we need to define the semantics of the new quantifiers and establish some of their logical properties. Although a formal definition would be premature at this time, we may attempt to specify some desired properties.

Let P = {x|p(x)} and Q = {x|q(x)} be discrete sets. Let |X| denote the cardinality of set X. Then we shall consider the formula $\overset{\circ}{\forall}\{x\in P\}[q(x)]$ as being true [13] iff the ratio $\frac{|P\cap Q|}{|P|}$ approaches 1 when |P| is increased; otherwise the formula is false. Similarly, we shall consider the formula $\overset{\circ}{\exists}\{x\in P\}[q(x)]$ as true iff the ratio $\frac{|P\cap Q|}{|P|}$

[11] Obviously, the evidence we have is only finite but the statement seems to go beyond this.

[12] In general, a coordinate represents decompositions more accurately. For example, it may be more appropriate to decompose physicists into individuals and groups who won a prize.

[13] with respect to an interpretation

does not approach 0 when |P| is increased; otherwise the formula is false. The above could be restated formally in terms of limits, but there seem to be several problems with such interpretation. For one thing, generic interpretations could not be assigned to sentences addressing large but finite sets (*NYU undergraduates like pizza*). On the other hand, we do not have to take limits literally, but merely as a convenient way of expressing a certain noticeable general *pattern*, as we examine given evidence, piece by piece. Doubtless, we need a sizable amount of data to observe any general pattern, but then we jump to conclusions which may extend over all unexamined instances *no matter how many of them are there*. Another problem is that the ratio of 1 may be too stringent a criterion for the truth of a transparently generic statement. Some people feel that generics are *fuzzier* than that. This may have to do with the problem of finite sets, or with logical properties of generics more generally.

These problems notwithstanding, a definition along the lines suggested above would make the new quantifiers are duals with respect to negation (Barwise & Cooper 1981), which is intuitively correct.

$$\neg \overset{\circ}{\forall} \{x \in P\}[q(x)] \text{ iff } \overset{\circ}{\exists} \{x \in P\}[\neg q(x)]$$

$$\neg \overset{\circ}{\exists} \{x \in P\}[q(x)] \text{ iff } \overset{\circ}{\forall} \{x \in P\}[\neg q(x)]$$

Thus, *birds fly* is false in its transparent reading just in the case when the number of birds that do not fly is non-negligible among all birds. Note that the negation of *birds fly* is a weaker statement than *birds do not fly*. Indeed, *birds do not fly* implies falsehood of *birds fly*, but not vice versa. Similarly, *Physicists win Nobel Prizes* is false in its opaque reading when among all Nobel Prize winners all but a negligible number are not physicists.

In the conclusion, let us return to the problem of representation for generic sentences. Let P and Q be sets defined as before. Let further $C_P \subseteq P$ and $C_Q \subseteq Q$ be reference contexts imposed on sets P and Q, respectively. Let P' and Q' be the translations of the generic terms referring to P and Q respectively (note that by this we also assume the decomposition coordinate). Then the generic sentence with the higher-level translation Q'(P') can be given one or both of the following lower-level interpretations:

(1) transparent interpretation: $\overset{\circ}{\forall} \{x \in P \cap C_P\}[q(x)]$

(2) opaque interpretation: $\overset{\circ}{\exists} \{x \in Q \cap C_Q\}[p(x)]$

The transparent/opaque ambiguity of a generic sentence can be often cleared away if bare plural is replaced by a phrase with a determiner such as *many, often, usually, some, hardly ever*, perhaps even *most*, etc. Thus, *most physicists win Nobel Prizes* appears to have only the transparently generic reading, while *birds often fly* seems to shift the preferred reading on the opaquely generic side. When the context is properly bound *most* may translate as *nearly all*, while *many* and *often* may be taken as *non-negligible amount*.

5. Conclusion

This paper presents a new look at the problem of finding a satisfactory meaning representation for generic sentences in natural language. Although the presentation is mostly informal and intuitive, some effort has been made to put the intuitions into an organized, semi-formal setting. The paper addresses only a limited range of problems, concentrating largely on two possible interpretations of generic sentences: transparent and opaque. Other possible types of interpretations for non-singular terms and sentences, including habitual statements, are not discussed here. We also ignore the problem of interaction between various readings of a non-singular sentence that might affect its final interpretation.

The paper suggests an extension to first-order representation languages that are necessary to capture the notion of non-singularity and genericity. At this time, it is not known how this extension would affect the computability of a knowledge representation language, specifically the tractability of an inference system. We point out that default and nonmonotonic logics based on the notions of domain closure or predicate circumscription, appear inadequate for representing meaning of generic statements.

6. Acknowledgments

The research reported here benefited greatly from the discussions with the members of the Natural Language Group: Mahesh Chitrao, Carol Friedman, Ralph Grishman, Tomasz Ksiezyk, Ping Peng, Sashidhar Reddi, John Sterling and Xu Zhiman. This report is based upon work supported by the Defense Advanced Research Project Agency under Contract N00014-85-K-0163 from the Office of Naval Research.

7. References

Barwise, J., R. Cooper (1981). "Generalized quantifiers and natural language." *Linguistics and Philosophy*. vol. 4, no. 2. pp. 159-217.

Carlson, G. N. (1982). "Generic Terms and Generic Sentences." **Journal of Philosophical Logic**. 11 pp. 145-181.

Kripke, S. (1972). "Naming and Necessity." In D. Davison and G. Harman (eds.), *Semantics of Natural Language*. Reidel, Dordrecht. pp. 253-355.

McCarthy, J. (1980). "Circumscription - A form of nonmonotonic reasoning." *Artificial Intelligence*, 13(1-2), pp. 27-39.

Montague, R. (1974) *Formal Philosophy. Selected Papers of Richard Montague*. R. H. Thomason (ed), Yale University Press.

Partee, B. H. (1978). "Bound variables and other anaphors." In D. L. Waltz (ed.), *TINLAP-2: Theoretical Issues in natural language processing*. pp. 79-85.

Quine, W. V. (1960). *Word and Object*. The M.I.T. Press. Cambridge, Mass.

Reiter, R. (1980). "A logic for default reasoning." *Artificial Intelligence*, 13(1-2), pp. 81-132.

Scha, R., D. Stallard (1988). "Multi-level plurals and distributivity." In Proceedings of the 26th Meeting of the ACL, Buffalo, NY. pp. 17-24.

Schubert, L. K., F. J. Pelletier (1987). "Problems in the Representation of the Logical Form of Generics." In E. LePore (ed.), *New Directions in Sementics*. Academic Press, London, pp. 385-451.

Sidner, C. L. (1979). "The role of focusing in interpretation of pronouns." Proceedings of the *17th Annual Meeting of the ACL*. pp. 77-78.

Strzalkowski, T. (1986a). "Representing Contextual Dependencies in Discourse." *Proceedings: Canadian Conference on Artificial Intelligence (CSCSI/SCEIO'86)*. Montreal.

Strzalkowski, T. (1986b). "An Approach to Non-Singular Terms in Discourse." *Proceedings: 11th Int. Conference on Computational Linguistics (COLING)*. Bonn, Germany.

Strzalkowski, T. (1986c). *A Theory of Stratified Meaning Representation*. Doctoral Dissertation. Simon Fraser University, Burnaby, B.C.

Strzalkowski, T., N. Cercone (1986). *A Framework for Computing Extra-sentential References*. **Computational Intelligence**. 2(4). pp. 159-179.

Strzalkowski, T., N. Cercone (1988). "Non-singular Terms in Natural Language Discourse." Submitted to **Computational Linguistics**.

Strzalkowski, T. (1988). "A logic based approach to discourse analysis." In preparation.

NYU COMPSCI TR-423
Strzalkowski, Tomek
A meaning representation
for generic sentences.
c.1

NYU COMPSCI TR-423
Strzalkowski, Tomek
A meaning representation
for generic sentences.
c.1

DATE DUE	BORROWER'S NAME

This book may be kept

FOURTEEN DAYS

A fine will be charged for each day the book is kept overtime.

GAYLORD 142			PRINTED IN U S A